Computers and Mathematics with Applications 57 (2009) 202–216

Contents lists available at ScienceDirect

Computers and Mathematics with Applications

journal homepage: www.elsevier.com/locate/camwa

The information content of fuzzy relations and fuzzy rules☆

Dan Hu[a,*], Hongxing Li[b], Xianchuan Yu[a]

[a] College of Information Science and Technology, Beijing Normal University, Beijing 100875, PR China

[b] School of Electronic and Information Engineering, Dalian University of Technology, Dalian, 116024, PR China

ARTICLE INFO

Article history:
Received 9 February 2007
Received in revised form 30 September 2008
Accepted 16 October 2008

Keywords:
Fuzzy relations
Fuzzy rules
Information content of fuzzy relations
Information content of fuzzy rules
Implication operator

ABSTRACT

Based on different emphases, some general information content measures of fuzzy relations are given. Distinguished from the measures related to fuzzy relations proposed before, these measures try to estimate the information conveyed by multiple-domain fuzzy relations without preassigned probability distribution. Since fuzzy rules can be fully captured by fuzzy relations from an input universe to an output universe, the information content of fuzzy rules can be easily measured by the information content of fuzzy relations proposed by us. However, there also exists a difference between fuzzy relations and fuzzy rules. Rules(especially classification rules) always have a direction from the antecedent and consequent while relations do not have direction. Based on this difference, some general measures for the information content of fuzzy rules are proposed. In practice, these measures can do well in the evaluation and selection of fuzzy rules. Finally, the measures of the information content of fuzzy rules are used to evaluate the stability and sensitivity of fuzzy implication operators in fuzzy control.

1. Introduction

Fuzzy relations, as an important concept in fuzzy theory, has been widely used in many fields such as fuzzy clustering, uncertainty reasoning and fuzzy control. When fuzzy relations are used in practice, how to estimate and compare them is a significant problem. Some related researches have been done to measure the uncertainty of fuzzy relations. Yager introduced an uncertainty measure for similarity relations and discussed its application to questionnaire design [1]. Hernandez and Recasens extended Yager's work and presented the formulae of joint entropy and conditional entropy based on the measure, and used the measures to learn fuzzy decision trees [2]. Yu presented some general uncertainty measures for fuzzy binary relations [3] and used them to define the diversity of multiple classifiers systems and granularity of granular computing. Furthermore, the literature on fuzzy relations and fuzzy partitions which correspond to fuzzy equivalent relations are abundant. Based on Ruspini's definition of fuzzy partitions [4] and the probability of fuzzy event defined by Zadeh [5], Tanaka put forward the entropy of fuzzy partition [6]. Dumitrescu extended the measure proposed by Tanaka by fuzzy measure, T-norms and a new definition of fuzzy partition [7], and this definition was then generalized by Mesiar [8]. Based on aggregation operators and the crisp measures of α-cuts, the uncertainty measure on fuzzy ε-partitions was proposed by Bertoluzza [9], and this measure is one of the few measures which are not associated with probability distribution. There is also other literature related to the evaluation of fuzzy partitions [10,11], which will not be stated in detail because they did not focus on the emphasis of this paper: entropy and information content.

☆ Supported by The Nature Science Foundation of China (Grant No. 60774049; No. 40672195); Program for New Century Excellent Talents in University (No. NCET-06-0131); the National 973 Fundamental Research Project of China (No. 2002CB312200); the National 863 High-Tech Program of China (No. 2006AA04Z163).

* Corresponding author.
E-mail addresses: hufengdd@163.com (D. Hu), lhxqx@bnu.edu.cn (H.X. Li), yuxianchuan@163.com (X.C. Yu).

0898-1221/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2008.10.087

When we discuss the estimation methods of fuzzy relations, we must make sure which characteristics the measures focus on:

Entropy or information content: entropy measures how much uncertainty exists in fuzzy relations, while information content measures the amount of information conveyed by the fuzzy relations. The information content of a fuzzy relation is the difference between the entropy before and after the fuzzy relation is obtained.

Probabilistic or non-probabilistic: a measure is probabilistic if it is based on a preassigned probability distribution, while the non-probabilistic measures deal with fuzzy relations without probability distribution.

One domain or multiple-domain: similarity and equivalence relations are all defined on a single domain, while general relations are constructed on two or more different domains.

In the past literature related to fuzzy relations [1–9], most of them emphasized entropy, and put forward measures of fuzzy relations defined on a single domain with probability distribution preassigned. But the measurement of information content of fuzzy relations defined on multiple-domains without preassigned probability distribution is needed. For example, experts give us some fuzzy rules without probability distribution of domains, and we want to compare them and choose several fuzzy rules with more information by the discussion of the multiple-domain fuzzy relations related with them. So, the problem we want to solve in this paper is how to measure the information conveyed by multiple-domain fuzzy relations without the preassigned probability distribution.

The paper is organized as follows: in Section 2, we present the basic measures and properties of the information content of fuzzy relations on discrete domains and continuous domains; in Section 3, the information content of fuzzy rules and fuzzy rule bases are discussed, and seven useful measures are given. Two examples show how to use the new measures and how to compare different fuzzy implication operators by them. Finally, the conclusion is given in Section 4.

2. The information content of fuzzy relations

2.1. The information content of fuzzy relations defined on discrete domains

A crisp relation represents the presence or absence of association, interaction, or interconnectedness between the elements of two or more sets [12]. A relation among crisp sets $X_1, X_2, \ldots, X_n$ is a subset of the Cartesian product $\prod_{i=1}^{n} X_i$. It is denoted either by $R(X_1, X_2, \ldots, X_n)$ or by the abbreviated form R. Each crisp relation R can be defined by a characteristic function that assigns a value of 1 to every tuple of the universal set belonging in the relation and a 0 to every tuple that does not belong. Thus, a crisp relation takes values in $\{0, 1\}$,

$$\mu_R(x_1, x_2, \ldots, x_n) = R(x_1, x_2, \ldots, x_n) = \begin{cases} 1 & \text{if and only if } (x_1, x_2, \ldots, x_n) \in R \\ 0 & \text{otherwise.} \end{cases}$$

In fuzzy set theory, degrees of association between elements can be represented by membership grades in a fuzzy relation in the same way as degrees of set membership are represented in the fuzzy set. Fuzzy relations take values in the interval $[0, 1]$.

A fuzzy relation $R(X, X)$ is reflexive if and only if $R(x, x) = 1 (\forall x \in X)$; $R(X, X)$ is symmetric if and only if $R(x, y) = R(y, x) (\forall x, y \in X)$; $R(X, X)$ is T-transitivity if and only if $T(R(x, y), R(y, z)) \le R(x, z) (\forall x, y, z \in X)$, where T is a t-norm. $R(X, X)$ is called a fuzzy similarity relation if it is reflexive and symmetric. If $R(X, X)$ is a fuzzy similarity relation and satisfies the property of T-transitivity, $R(X, X)$ is a T-indistinguishability relation or fuzzy equivalence relation.

At first, the information content of a binary fuzzy relation will be defined according to the interconnection between the given universes X and Y.

Definition 1. Let $X = \{x_1, x_2, \ldots, x_n\}$, $Y = \{y_1, y_2, \ldots, y_m\}$, R is a fuzzy relation from X to Y, $R \subseteq X \times Y$. The inverse fuzzy relation R^{-1} is defined by

$$R^{-1}(y, x) = R(x, y).$$

The domain and range of R is also a fuzzy set and the membership functions of them is

$$\mathrm{dom}R(x) = \max_{y \in Y} R(x, y),$$
$$\mathrm{ran}R(y) = \max_{x \in X} R(x, y).$$

The image of $x_i \in X$ and the inverse image of $y_j \in Y$ are also fuzzy sets and defined as:

$$R(x_i)(y) = R(x_i, y),$$
$$R^{-1}(y_j)(x) = R(x, y_j).$$

(1) When $\mathrm{dom}R = X$, $\mathrm{ran}R = Y$, The information content of R is measured as follows:

$$IC_1(R) = \frac{n}{m+n} IC_1(R \Downarrow X) + \frac{m}{m+n} IC_1(R \Downarrow Y), \tag{1}$$

where $IC_1(R \Downarrow X)$ and $IC_1(R \Downarrow Y)$ are the information content of R restricted on X and Y respectively,

$$IC_1(R \Downarrow X) = -\sum_{i=1}^{n} \frac{\sum_{j=1}^{m} R(x_i, y_j)}{\sum_{i=1}^{n}\sum_{j=1}^{m} R(x_i, y_j)} \log \frac{\sum_{j=1}^{m} R(x_i, y_j)}{m}. \tag{2}$$

$$IC_1(R \Downarrow Y) = -\sum_{j=1}^{m} \frac{\sum_{i=1}^{n} R^{-1}(y_j, x_i)}{\sum_{j=1}^{m}\sum_{i=1}^{n} R^{-1}(y_j, x_i)} \log \frac{\sum_{i=1}^{n} R^{-1}(y_j, x_i)}{n}. \tag{3}$$

(2) When $\operatorname{ran}R \subsetneq Y$ or $\operatorname{dom}R \subsetneq X$, let

$$R'(x_i)(y) = \begin{cases} R(x_i)(y) & x_i \in \operatorname{dom}R \\ 1 & x_i \overline{\in} \operatorname{dom}R \end{cases} \quad (y \in Y) \tag{4}$$

$$(R'')^{-1}(y_j)(x) = \begin{cases} R^{-1}(y_j)(x) & y_j \in \operatorname{ran}R \\ 1 & y_j \overline{\in} \operatorname{ran}R \end{cases} \quad (x \in X) \tag{5}$$

The information content of R is defined as follows:

$$IC_1(R) = \frac{n}{m+n} IC_1(R' \Downarrow X) + \frac{m}{m+n} IC_1(R'' \Downarrow Y). \tag{6}$$

The base of logarithm is 2, and $0 \log 0 = 0$. The unit of information content is a "bit".

Remark 1. $IC_1(R \Downarrow X)$ and $IC_1(R \Downarrow Y)$ measure the average information of the projections from X to Y and the projections from Y to X respectively. That is, $IC_1(R \Downarrow X)$ and $IC_1(R \Downarrow Y)$ have directions. But in Definition 1, $IC_1(R)$ is the average of $IC_1(R \Downarrow X)$ and $IC_1(R \Downarrow Y)$ without direction, just because relation R describes the connection between X and Y and "connection" has no direction.

Remark 2. In Eq. (2),

$$-\log \frac{\sum_{j=1}^{m} R(x_i, y_j)}{m} \tag{7}$$

measures the information of the projection from x_i to Y. If x_i is not in $\operatorname{dom}R$, all elements in Y are assigned to x_i, which means there is no information conveyed by x_i. This is just the meaning of Eqs. (4) and (5).

A fuzzy relation is a fuzzy set indeed, so, the U-uncertainty of fuzzy set [12] can help us to find another way to measure the information content of fuzzy relations.

Definition 2 (*[12,13]*). A is a fuzzy set defined on $X = \{x_1, x_2, \ldots, x_n\}$. All $A(x_i)(i = 1, \ldots, n)$ can be designed to an ordered possibility distribution

$$\{\lambda_1, \lambda_2, \ldots, \lambda_n\}.$$

It is always the case that $\lambda_{i+1} \le \lambda_i$. The U-uncertainty of A can be defined as

$$U(A) = -\sum_{i=1}^{n} (\lambda_i - \lambda_{i+1}) \log |A_{\lambda_i}|, \tag{8}$$

where $\lambda_{n+1} = 0$ by convention, $|\cdot|$ is the cardinality of a set, and

$$A_{\lambda_i} = \{x \in X | A(x) \ge \lambda_i\}.$$

Definition 3. R is the fuzzy relation from X to Y defined in Definition 1.

All $R(x_i, y_j)(i = 1, \ldots, n; j = 1, \ldots, m)$ can be designed to an ordered possibility distribution

$$\{\lambda_1, \lambda_2, \ldots, \lambda_{m\times n}\}.$$

It is always the case that $\lambda_{i+1} \le \lambda_i$. The information content of R can be measured by $IC_2(R)$,

$$IC_2(R) = -\sum_{i=1}^{m\times n}(\lambda_i - \lambda_{i+1}) \log \frac{|R_{\lambda_i}|}{m \times n}, \tag{9}$$

where $\lambda_{n+1} = 0$ by convention.

In fact,

$$-\log \frac{|R_{\lambda_i}|}{m \times n} \triangleq ic_{\lambda_i}(R) \tag{10}$$

is the difference between the U-uncertainty of crisp sets $X \times Y$ and R_{λ_i}, thus Eq. (10) can denote the information of R at the level λ_i.

The λ-cut of fuzzy relation R is a crisp relation from X to Y. We can use the measure of the information content of crisp relations which we have defined in [14] to propose another measure of the information content of fuzzy relations.

Definition 4. All suppositions are same as Definition 3.

The information content of R can be measured by $IC_3(R)$,

$$IC_3(R) = \sum_{i=1}^{m\times n}(\lambda_i - \lambda_{i+1})H(R_{\lambda_i}), \tag{11}$$

where $\lambda_{n+1} = 0$ by convention, and $H(R_{\lambda_i})$ is the information content of R_{λ_i} defined in [14].

Based on the meaning of $ic_{\lambda_i}(R)$ and $H(R_{\lambda_i})$, the aggregation of the information of R at all levels λ_i can also be a measure of the information content of R. At this time, the idea of aggregation operators [15] can be easily used in the following definition.

Definition 5. R is the fuzzy relation from X to Y defined in Definition 1. Suppose there are q different values in $\{R(x_i, y_j)(i = 1, \dots, n; j = 1, \dots, m)\}$, and these q values can be designed to an ordered set

$$\{\mu_1, \mu_2, \dots, \mu_q\}.$$

It is always the case that $\mu_{i+1} < \mu_i$. The information content of R can be measured by $IC_4(R)$ or $IC_5(R)$,

$$IC_4(R) = -\sum_{i=1}^{q} \frac{\mu_i^\alpha}{\sum_{i=1}^{q}\mu_i^\alpha} \log \frac{|R_{\mu_i}|}{m \times n} = \sum_{i=1}^{q} \frac{\mu_i^\alpha}{\sum_{i=1}^{q}\mu_i^\alpha} ic_{\mu_i}(R), \quad (\alpha \ge 1) \tag{12}$$

$$IC_5(R) = \sum_{i=1}^{q} \frac{\mu_i^\alpha}{\sum_{i=1}^{q}\mu_i^\alpha} H(R_{\mu_i}), \quad (\alpha \ge 1). \tag{13}$$

If we use the aggregation operators defined in [15,16], the information content of fuzzy relation R can be defined as

$$IC_4^*(R) = G(ic_{\mu_1}(R), ic_{\mu_2}(R), \dots, ic_{\mu_q}(R)), \tag{14}$$

$$IC_5^*(R) = G(H(R_{\mu_1}), H(R_{\mu_2}), \dots, H(R_{\mu_q})), \tag{15}$$

where $G : \mathbb{R}^k \to \mathbb{R}$ is a aggregation operator.

$IC_4(R)$ and $IC_5(R)$ are specializations of $IC_4^*(R)$ and $IC_5^*(R)$ respectively.

Until now, five different measures have been proposed to scale the information content of fuzzy relation R, which emphasize different sides of R. We can choose them based on different needs in practice and their shared properties and differences are partially shown in following theorems.

Theorem 1 (*Maximum and Minimum*). *Let $X = \{x_1, x_2, \dots, x_n\}$, $Y = \{y_1, y_2, \dots, y_m\}$. R is a fuzzy relation from X to Y, $R \subseteq X \times Y$.*

$$0 \le IC_1(R) \le \frac{m}{m+n}\log n + \frac{n}{m+n}\log m.$$

$$0 \le IC_2(R) \le \log m \cdot n, \qquad 0 \le IC_4(R) \le \log m \cdot n,$$

$$0 \le IC_3(R) \le \frac{m}{m+n}\log n + \frac{n}{m+n}\log m,$$

$$0 \le IC_5(R) \le \frac{m}{m+n}\log n + \frac{n}{m+n}\log m.$$

Let $0 < \lambda_1 < 1$, R_1, R_2, R_3 and R_4 be relations from X to Y,

$$R_1 = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & 1 & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix} \qquad R_2 = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix},$$

then, when $R = R_1$, $IC_1(R)$, $IC_3(R)$ and $IC_5(R)$ reach their maximum; when $R = R_2$, $IC_2(R)$ and $IC_4(R)$ reach their maximum. Let

$$R_3 = \begin{pmatrix} \lambda_1 & \lambda_1 & \cdots & \lambda_1 \\ \lambda_1 & \lambda_1 & \cdots & \lambda_1 \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1 & \lambda_1 & \cdots & \lambda_1 \end{pmatrix} \qquad R_4 = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix},$$

$$IC_1(R_3) = \log\lambda_1, \qquad IC_2(R_3) = IC_3(R_3) = 0IC_4(R_3) = IC_5(R_3) = 0.$$
$$IC_1(R_4) = IC_2(R_4) = IC_3(R_4) = C_4(R_4) = IC_5(R_4) = 0.$$

Remark 3. Theorem 1 shows us $IC_1(R)$, $IC_2(R)$ and $IC_4(R)$ emphasize on the corresponding relation between X and Y. So, we we choose an element $x \in X$, the uncertainty in matching this element to some $y \in Y$ is $\log n$. When the corresponding element y is determined and the level is equal to 1, the uncertainty is completely reduced and the information gained is equal to $\log n$. Nevertheless, $IC_2(R)$ and $IC_4(R)$ take R as a fuzzy set and emphasize the distinguishability of elements $\{x_i, y_j\}$ in $X \times Y$. There are $n \times m$ elements needed to be determined in the relation matrix. Thus the uncertainty of R at this time is $\log m \cdot n$. When there is only one element in the relation R and the membership degree is equal to 1, the uncertainty is completely reduced.

Theorem 2 (*Branching Theory*). *Let $X = \{x_1, x_2, \ldots, x_n\}$, $Y = \{y_1, y_2, \ldots, y_m\}$, R_l, R_{lt2} and R_{lt3} $(l = 1, 2, 3\ t = 1, 2)$ are fuzzy relations from X to Y.*

(1) *All $R(x_i, y_j)$ can be designed to an ordered possibility distributions*

$$\{\lambda_1, \lambda_2, \ldots, \lambda_{m\times n}\}.$$

It is always the case that $\lambda_{i+1} \le \lambda_i$. Let

$$K1 = \{(x_i, y_j) \in X \times Y | R(x_i, y_j) = \lambda_{k+1}\},$$
$$K2 = \{(x_i, y_j) \in X \times Y | R(x_i, y_j) = \lambda_{k-1}\},$$
$$R_1(x_i, y_j) = \begin{cases} R(x_i, y_j) & (x_i, y_j) \overline{\in} K1 \\ \lambda_k & (x_i, y_j) \in K1 \end{cases}$$
$$R_{112}(x_i, y_j) = \begin{cases} 1 & R(x_i, y_j) \ge \lambda_k \\ \dfrac{\lambda_{k-1} - \lambda_{k+2}}{\lambda_k - \lambda_{k+2}} & R(x_i, y_j) = \lambda_{k+1} \\ 0 & R(x_i, y_j) < \lambda_{k+1} \end{cases}$$
$$R_{113}(x_i, y_j) = \begin{cases} 1 & R(x_i, y_j) \ge \lambda_{k+1} \\ 0 & R(x_i, y_j) < \lambda_{k+1} \end{cases}$$
$$R_2(x_i, y_j) = \begin{cases} R(x_i, y_j) & (x_i, y_j) \overline{\in} K2 \\ \lambda_k & (x_i, y_j) \in K2 \end{cases}$$
$$R_{212}(x_i, y_j) = \begin{cases} 1 & R(x_i, y_j) \ge \lambda_{k-2} \\ \dfrac{\lambda_{k-1} - \lambda_k}{\lambda_{k-2} - \lambda_k} & R(x_i, y_j) = \lambda_{k-1} \\ 0 & R(x_i, y_j) < \lambda_{k-1} \end{cases}$$
$$R_{213}(x_i, y_j) = \begin{cases} 1 & R(x_i, y_j) \ge \lambda_{k-2} \\ 0 & R(x_i, y_j) < \lambda_{k-2}. \end{cases}$$

Then,

$$IC_2(R) = IC_2(R_1) + (\lambda_k - \lambda_{k+2}) IC_2(R_{112}) - (\lambda_k - \lambda_{k+2}) IC_2(R_{113}).$$
$$IC_3(R) = IC_3(R_1) + (\lambda_k - \lambda_{k+2}) IC_3(R_{112}) - (\lambda_k - \lambda_{k+2}) IC_3(R_{113}).$$
$$IC_2(R) = IC_2(R_2) + (\lambda_{k-2} - \lambda_k) IC_2(R_{212}) - (\lambda_{k-2} - \lambda_k) IC_2(R_{213}).$$
$$IC_3(R) = IC_3(R_2) + (\lambda_{k-2} - \lambda_k) IC_3(R_{212}) - (\lambda_{k-2} - \lambda_k) IC_3(R_{213}).$$

(2) *Suppose there are q different values in* $\{R(x_i, y_j)(i = 1, \dots, n; j = 1, \dots, m)\}$, *and these q values can be designed to an ordered set*

$$\{\mu_1, \mu_2, \dots, \mu_q\}.$$

It is always the case that $\mu_{i+1} < \mu_i$. *Let*

$$R_{122}(x_i, y_j) = \begin{cases} \lambda_k & R(x_i, y_j) \geq \lambda_k \\ \lambda_{k+1} & R(x_i, y_j) = \lambda_{k+1} \\ 0 & R(x_i, y_j) \leq \lambda_{k+2} \end{cases}$$

$$R_{123}(x_i, y_j) = \begin{cases} 1 & R(x_i, y_j) \geq \lambda_{k+1} \\ 0 & R(x_i, y_j) < \lambda_{k+1} \end{cases}$$

$$R_{222}(x_i, y_j) = \begin{cases} \lambda_{k-1} & R(x_i, y_j) \geq \lambda_{k-1} \\ \lambda_k & R(x_i, y_j) = \lambda_k \\ 0 & R(x_i, y_j) \leq \lambda_{k+1} \end{cases}$$

$$R_{223}(x_i, y_j) = \begin{cases} 1 & R(x_i, y_j) \geq \lambda_k \\ 0 & R(x_i, y_j) < \lambda_k. \end{cases}$$

Then,

$$IC_4(R) = \frac{-\mu_{k-1}^{\alpha} + \sum_{i=1}^{q} \mu_i^{\alpha}}{\sum_{i=1}^{q} \mu_i^{\alpha}} IC_4(R_2) + \frac{\mu_k^{\alpha} + \mu_{k-1}^{\alpha}}{\sum_{i=1}^{q} \mu_i^{\alpha}} IC_4(R_{222}) - \frac{\mu_k^{\alpha}}{\sum_{i=1}^{q} \mu_i^{\alpha}} IC_4(R_{223}).$$

$$IC_5(R) = \frac{-\mu_{k-1}^{\alpha} + \sum_{i=1}^{q} \mu_i^{\alpha}}{\sum_{i=1}^{q} \mu_i^{\alpha}} IC_5(R_2) + \frac{\mu_k^{\alpha} + \mu_{k-1}^{\alpha}}{\sum_{i=1}^{q} \mu_i^{\alpha}} IC_5(R_{222}) - \frac{\mu_k^{\alpha}}{\sum_{i=1}^{q} \mu_i^{\alpha}} IC_5(R_{223}).$$

$$IC_4(R) = \frac{-\mu_{k+1}^{\alpha} + \sum_{i=1}^{q} \mu_i^{\alpha}}{\sum_{i=1}^{q} \mu_i^{\alpha}} IC_4(R_1) + \frac{\mu_k^{\alpha} + \mu_{k+1}^{\alpha}}{\sum_{i=1}^{q} \mu_i^{\alpha}} IC_4(R_{122}) - \frac{\mu_k^{\alpha}}{\sum_{i=1}^{q} \mu_i^{\alpha}} IC_4(R_{123}).$$

$$IC_5(R) = \frac{-\mu_{k+1}^{\alpha} + \sum_{i=1}^{q} \mu_i^{\alpha}}{\sum_{i=1}^{q} \mu_i^{\alpha}} IC_5(R_1) + \frac{\mu_k^{\alpha} + \mu_{k+1}^{\alpha}}{\sum_{i=1}^{q} \mu_i^{\alpha}} IC_5(R_{122}) - \frac{\mu_k^{\alpha}}{\sum_{i=1}^{q} \mu_i^{\alpha}} IC_5(R_{123}).$$

(3) $\forall x_0 \in \mathrm{dom}R$, $\forall y_0 \in \mathrm{ran}R$, *let*

$$R_{311}(x_i, y_j) = \begin{cases} R(x_i, y_j) & x_i \neq x_0 \\ 0 & x_i = x_0 \end{cases}$$

$$R_{312}(x_i, y_j) = \begin{cases} 0 & x_i \neq x_0 \\ R(x_i, y_j) & x_i = x_0 \end{cases}$$

$$R_{321}(x_i, y_j) = \begin{cases} R(x_i, y_j) & y_j \neq y_0 \\ 0 & y_j = x_0 \end{cases}$$

$$R_{322}(x_i, y_j) = \begin{cases} 0 & y_j \neq y_0 \\ R(x_i, y_j) & y_j = y_0. \end{cases}$$

Then,

$$IC_1(R \Downarrow X) = \frac{-m + \sum_{i=1}^{n} \sum_{j=1}^{m} R(x_i, y_j)}{\sum_{i=1}^{n} \sum_{j=1}^{m} R(x_i, y_j)} IC_1(R_{311} \Downarrow X) + \frac{(n-1)m + \sum_{j=1}^{n} R(x_0, y_j)}{\sum_{i=1}^{n} \sum_{j=1}^{m} R(x_i, y_j)} IC_1(R_{312} \Downarrow X),$$

$$IC_1(R \Downarrow Y) = \frac{-n + \sum_{i=1}^{n}\sum_{j=1}^{m} R(x_i, y_j)}{\sum_{i=1}^{n}\sum_{j=1}^{m} R(x_i, y_j)} IC_1(R_{321} \Downarrow Y) + \frac{(m-1)n + \sum_{i=1}^{n} R(x_i, y_0)}{\sum_{i=1}^{n}\sum_{j=1}^{m} R(x_i, y_j)} IC_1(R_{322} \Downarrow Y),$$

Proof. We prove the equation

$$IC_3(R) = IC_3(R_1) + (\lambda_k - \lambda_{k+2}) IC_3(R_{122}) - (\lambda_k - \lambda_{k+2}) IC_3(R_{132}).$$

$$\begin{aligned} IC_3(R_1) &= \sum_{i=1}^{k-1} (\lambda_i - \lambda_{i+1}) H((R_{11})_{\lambda_i}) + (\lambda_k - \lambda_{k+2}) H((R_{11})_{\lambda_k}) + \sum_{i=k+2}^{m \times n} (\lambda_i - \lambda_{i+1}) H((R_{11})_{\lambda_i}) \\ &= \sum_{i=1}^{k-1} (\lambda_i - \lambda_{i+1}) H(R_{\lambda_i}) + (\lambda_k - \lambda_{k+2}) H(R_{\lambda_{k+1}}) + \sum_{i=k+2}^{m \times n} (\lambda_i - \lambda_{i+1}) H(R_{\lambda_i}) \end{aligned}$$

$$\begin{aligned} IC_3(R_{122}) &= \left(1 - \frac{\lambda_{k-1} - \lambda_{k+2}}{\lambda_k - \lambda_{k+2}}\right) H((R_{122})_1) + \frac{\lambda_{k-1} - \lambda_{k+2}}{\lambda_k - \lambda_{k+2}} H((R_{122})_{\frac{\lambda_{k-1}-\lambda_{k+2}}{\lambda_k - \lambda_{k+2}}}) \\ &= \frac{\lambda_k - \lambda_{k-1}}{\lambda_k - \lambda_{k+2}} H(R_{\lambda_k}) + \frac{\lambda_{k-1} - \lambda_{k+2}}{\lambda_k - \lambda_{k+2}} H(R_{\lambda_{k+1}}) \end{aligned}$$

$$IC_3(R_{132}) = (1 - 0) H((R_{132})_1) = H(R_{\lambda_{k+1}})$$

$$\begin{aligned} & IC_3(R_{11}) + (\lambda_k - \lambda_{k+2}) IC_3(R_{122}) - (\lambda_k - \lambda_{k+2}) IC_3(R_{132}) \\ &= \sum_{i=1}^{k-1} (\lambda_i - \lambda_{i+1}) H(R_{\lambda_i}) + (\lambda_k - \lambda_{k+2}) H(R_{\lambda_{k+1}}) + \sum_{i=k+2}^{m \times n} (\lambda_i - \lambda_{i+1}) H(R_{\lambda_i}) \\ &\quad + (\lambda_k - \lambda_{k-1}) H(R_{\lambda_k}) + (\lambda_{k-1} - \lambda_{k+2}) H(R_{\lambda_{k+1}}) - (\lambda_k - \lambda_{k+2}) H(R_{\lambda_{k+1}}) \\ &= \sum_{i=1}^{m \times n} (\lambda_i - \lambda_{i+1}) H(R_{\lambda_i}) = IC_3(R). \end{aligned}$$

The other part of the proof is same to it. □

Branching theory requires the information content measures to be capable of measuring the information in two ways. That is, the information content is measured either directly for the given corresponding relation or indirectly by adding information associated with a combination of relations that reflect a two-stage measuring process. But it must be noticed that different measures may have different meanings of branching. At the first stage of branching, the distinction between the possibility values assigned to any two neighboring components is ignored for $IC_2(R)$ and $IC_3(R)$, while any level for distinction is ignored for $IC_4(R)$ and $IC_5(R)$. Meanwhile, for $IC_1(R)$, the corresponding relation of any element in universe is ignored at this stage.

Theorem 3 (*Information Symmetry*). *Let $X = \{x_1, x_2, \ldots, x_n\}$, $Y = \{y_1, y_2, \ldots, y_m\}$, R be a fuzzy relation from X to Y. R^{-1} is the inverse of R, then*

$$IC_t(R) = IC_t(R^{-1}), \quad t = 1, 2, 3, 4, 5.$$

2.2. The information content of fuzzy relations defined on continuous domains

A fuzzy relation defined on continuous domain can be seen as a hypersurface defined on the given universe. We can take the hypersurface as a whole and define the information content of fuzzy relations by the help of the volume of fuzzy sets [13].

Definition 6 (*[13]*). A is a fuzzy set defined on $\mathbb{X}$, $\mathbb{X}$ is a n-dimension space, $\mathbf{x} \in \mathbb{X}$. The volume of fuzzy set A is defined as

$$V(A) = \int_{\mathbb{X}} A(\mathbf{x}) \mathrm{d}\mathbf{x}. \tag{16}$$

Definition 7. R is a fuzzy relation defined on $\mathbb{X}$, the information content of R can be defined as

$$IC_6(R) = -\log \frac{V(R)}{\int_{\mathbb{X}} \mathrm{d}\mathbf{x}}. \tag{17}$$

By the definition of U-uncertainty of fuzzy sets on continuous domain [13], Definition 3 can be easily extended to the information content of fuzzy relations defined on continuous domains.

Definition 8. R is a fuzzy relation defined on $\mathbb{X}$, $\bar{R}(\mathbf{x})$ is the non-increasing rearrangement of $R(\mathbf{x})$ [13]. The information content of R can be defined as

$$IC_7(R) = \left(\frac{1}{\ln 2}\right) \int_{\mathbb{X}} \frac{1 - \bar{R}(\mathbf{x})}{\mathbf{x}} \mathrm{d}\mathbf{x}. \tag{18}$$

2.3. The information content of n-ary fuzzy relations

Consider the Cartesian product of all sets in the family $X = \{X_i | i \in I\}$. For each n-tuple

$$\mathbf{x} = (x_i | i \in I) \in \prod_{i \in I} X_i$$

and each r-tuple ($r < n$),

$$\mathbf{y} = (y_j | i \in J \subset I) \in \prod_{j \in J} X_i,$$

let $\mathbf{y}$ be called a subsequence of $\mathbf{x}$ if and only if $y_j = x_j$ for all $j \in J$. Let $\mathbf{y} \prec \mathbf{x}$ or $\mathbf{x} \succ \mathbf{y}$ denote that $\mathbf{y}$ is a subsequence of $\mathbf{x}$.

Given a relation $R(X_1, X_2, \ldots, X_n)$, let $[R \downarrow Y]$ denote the projection of R that disregards all variables in X except those in the set

$$Y = \{X_j | j \in J \subset I\}.$$

Then, $[R \downarrow Y]$ is a fuzzy set (relation) whose membership function is defined on the Cartesian product of sets in Y by the equation

$$[R \downarrow Y](\mathbf{y}) = \max_{\mathbf{x} \succ \mathbf{y}} R(\mathbf{x}),$$

where $R(\mathbf{x})$ is the membership function of the given n-ary relation R.

Definition 9. Let $X_i = \{x_{i1}, x_{i2}, \ldots, x_{im_i}\}$, $i \in I = \{1, \ldots, n\}$, and R be a n-ary fuzzy relation on $\prod_{i=1}^{n} X_i$.

(1) When $\forall i \in I$, $\mathrm{Supp}([R \downarrow X_i]) = X_i$, the information content of R is defined as follows:

$$IC_1(R) = \sum_{i=1}^{n} \frac{|X_i|}{\sum_{i=1}^{n} |X_i|} IC_1(R \Downarrow X_i) = \sum_{i=1}^{n} \frac{m_i}{\prod_{i=1}^{n} m_i} IC_1(R \Downarrow X_i), \tag{19}$$

$$IC_1(R \Downarrow X_i) = -\sum_{j=1}^{m_i} \frac{\sum_{\mathbf{x} \succ x_{ij}} R(\mathbf{x})}{\sum_{j=1}^{m_i} \sum_{\mathbf{x} \succ x_{ij}} R(\mathbf{x})} \log \frac{\sum_{\mathbf{x} \succ x_{ij}} R(\mathbf{x})}{\prod_{k \in I \setminus \{i\}} |X_k|}. \tag{20}$$

(2) When $\exists i_0 \in I$, $\mathrm{Supp}([R \downarrow X_{i_0}]) \subsetneq X_{i_0}$ let

$$R^{i_0}(\mathbf{x}) = \begin{cases} 1 & [R \downarrow X_{i_0}](x_{i_0 j}) = 0 \text{ and } \mathbf{x} \succ x_{i_0 j} \ (j = 1, \ldots, m_{i_0}) \\ R(\mathbf{x}) & \text{others.} \end{cases} \tag{21}$$

The information content of R is defined as follows:

$$IC_1(R \Downarrow X_{i_0}) \triangleq IC_1(R^{i_0} \Downarrow X_{i_0}), \tag{22}$$

$$IC_1(R) = \sum_{i=1}^{n} \frac{|X_i|}{\sum_{i=1}^{n} |X_i|} IC_1(R \Downarrow X_i), \tag{23}$$

where $\mathrm{Supp}(A)$ is the support set of A.

$IC_2(R)$ and $IC_4(R)$ in Definitions 3 and 5 can be easily generalized to n-ary fuzzy relation R by changing the item $m \times n$ in Eqs. (9) and (12) to $\prod_{k \in I} |X_k|$.

By the definition of the information content of crisp n-ary relations, which is the specialization of Definition 8, $IC_3(R)$ and $IC_5(R)$ in Definitions 4 and 5 can be generalized to n-ary fuzzy relation R.

Finally, $IC_6(R)$ and $IC_7(R)$ are defined on n-dimension space and can be applied to n-ary fuzzy relation R directly.

3. The information content of fuzzy rules and the comparison of implication operators

The use of rule bases is common in fuzzy models, fuzzy controllers and fuzzy expert systems. Furthermore, fuzzy rule bases and fuzzy inference are core parts of these intelligent systems. So, the discussion about fuzzy rules must bring some benefits to the progress of fuzzy modeling, fuzzy control and fuzzy expert systems. For example, fuzzy rule selection is one

Table 1
Some different implication operators ($A \to B$).

O^i	R^i	$R^i(x,y)$
O^1	$(A^c \times Y) \cup (A \times B)$	$(1 - A(x)) \vee (A(x) \wedge B(y))$
O^2	$(A^c \times Y) \cup (X \times B)$	$(1 - A(x)) \vee B(y)$
O^3	$A^c \oplus B$	$(1 - A(x) + B(y)) \wedge 1$
O^4	$A^c \uplus B$	$1 - A(x) + A(x) \cdot B(y)$
O^5	$A \cap B$	$A(x) \wedge B(y)$
O^6	$A \sqcup B$	$0 \vee (A(x) + B(y) - 1)$
O^7	$A \odot B$	$A(x) \cdot B(y)$

of the most important parts in fuzzy rule mining. Much work has been done to introduce useful measures for the evaluation of fuzzy rule. In these measures, confidence, support [17,18] and volume [19–22] are more frequently used in practice.

The confidence and support of a fuzzy rule "if A, then B", which is induced from database D, are given as follows [17]:

$$\mathrm{supp}(A \to B) = \sum_{(x,y)\in D} A(x) \otimes B(y), \tag{24}$$

$$\mathrm{conf}(A \to B) = \frac{\sum\limits_{(x,y)\in D} A(x) \otimes B(y)}{\sum\limits_{(x,y)\in D} A(x)}, \tag{25}$$

where $\otimes$ is a t-norm and the usual choice is $\otimes = \min$. It is obvious that support and confidence are measures closely related to the database D. When we want to evaluate fuzzy rules without the database or the fuzzy rules are not induced from a database but human experts or others, we must find some new way. The volume of fuzzy rules [19] may be a good measure, but it is only suitable for a fuzzy classification rule whose consequent is class label not fuzzy set. And it is very difficult to compute when there are more than two conditions in the antecedent of the rules. So, it is better to find some new measures to evaluate fuzzy rules without a database, and this is just the work we want to do in this section.

Let $\mathbb{K}$ be a fuzzy rule base, $\forall r \in \mathbb{K}$, the fuzzy rule r takes the form:

$$r: \quad \text{If } A_1 \text{and} A_2, \text{ and } \ldots, \text{ and } A_n, \text{ Then } B$$

where $A_1, A_2, \ldots, A_n$ and B are fuzzy sets defined on universal sets $X_1, X_2, \ldots, X_n$ of inputs and Y of outputs, respectively. The fuzzy rule r means that $A_1 \times A_2 \times \cdots \times A_n$ implies B, that is

$$A_1 \times A_2 \times \cdots \times A_n \to B.$$

The implication operator "$\to$" can be interpreted in many different ways, and different operators will bring different fuzzy relations on $\prod_{i=1}^n X_i \times Y$. Some different implication operators are shown in Table 1.

No matter what the implication operator is chosen, there are fuzzy relations R_r and $R_{\mathbb{K}}$ corresponding to the fuzzy rule r and rule base $\mathbb{K}$ respectively, where

$$R_{\mathbb{K}} = \bigcup_{r\in\mathbb{K}} R_r.$$

That is, a single rule r or the rule base $\mathbb{K}$ composed of rules can be fully captured by fuzzy relations defined on $\prod_{i=1}^n X_i \times Y$. Thus, we can use the information content of fuzzy relations to measure how much information the given rules have conveyed to us.

Definition 10. Let $\mathbb{K}$ be a rule base, $X_i = \{x_{i1}, x_{i2}, \ldots, x_{im_i}\}$, $i \in I = \{1, \ldots, n\}$. $\prod_{i=1}^n X_i$ and $Y = \{y_1, y_2, \ldots, y_m\}$ are universal sets of inputs and outputs. $R_{\mathbb{K}}$ is the fuzzy relation corresponding to $\mathbb{K}$. Suppose there are q different values in

$$\left\{ R(\mathbf{x}, y) | \mathbf{x} \in \prod_{i=1}^n X_i, y \in Y \right\},$$

and these q values can be designed to an ordered set

$$\{\mu_1, \mu_2, \ldots, \mu_q\}.$$

It is always the case that $\mu_{i+1} < \mu_i$. At the same time, all elements in $\{R(\mathbf{x}, y)\}$ can be designed to an ordered possibility distributions

$$\{\lambda_1, \lambda_2, \ldots, \lambda_{m\times\prod_{i=1}^n m_i}\}.$$

It is always the case that $\lambda_{i+1} \le \lambda_i$ and $\lambda_{1+m\times\prod_{i=1}^n m_i} = 0$.

The information content of rule base $\mathbb{K}$ can be defined as follows:

$$ICR_1(\mathbb{K}) = IC_1\left(R_{\mathbb{K}} \Downarrow \prod_{i=1}^{n} X_i\right), \tag{26}$$

$$ICR_2(\mathbb{K}) = IC_2(R_{\mathbb{K}}) = -\sum_{i=1}^{m\times\prod_{j=1}^{n} m_j} (\lambda_i - \lambda_{i+1}) \log \frac{|(R_{\mathbb{K}})_{\lambda_i}|}{m \times \prod_{j=1}^{n} m_j}, \tag{27}$$

$$ICR_3(\mathbb{K}) = \sum_{i=1}^{m\times\prod_{j=1}^{n} m_j} (\lambda_i - \lambda_{i+1}) H\left((R_{\mathbb{K}})_{\lambda_i} \Downarrow \prod_{s=1}^{n} X_s\right), \tag{28}$$

$$ICR_4(\mathbb{K}) = IC_4(R_{\mathbb{K}}) = -\sum_{i=1}^{q} \frac{\mu_i^{\alpha}}{\sum_{i=1}^{q} \mu_i^{\alpha}} \log \frac{|(R_{\mathbb{K}})_{\mu_i}|}{m \times \prod_{j=1}^{n} m_j}, \quad (\alpha \geq 1), \tag{29}$$

$$ICR_5(\mathbb{K}) = \sum_{i=1}^{q} \frac{\mu_i^{\alpha}}{\sum_{i=1}^{q} \mu_i^{\alpha}} H\left((R_{\mathbb{K}})_{\mu_i} \Downarrow \prod_{s=1}^{n} X_s\right), \quad (\alpha \geq 1), \tag{30}$$

where $H((R_{\mathbb{K}})_{\mu_i} \Downarrow \prod_{s=1}^{n} X_s)$ is the entropy of the crisp relation $(R_{\mathbb{K}})_{\mu_i}$ restricted on $\prod_{s=1}^{n} X_s$ which has been defined in [14].

Definition 11. Let $\mathbb{K}$ be a rule base, $\mathbb{X} = \prod_{i=1}^{n} X_i$ and Y are all continuous domains. $R_{\mathbb{K}}$ is the fuzzy relation defined from $X = \prod_{i=1}^{n} X_i$ to Y and corresponds to the fuzzy rule base $\mathbb{K}$. The information content of rule base $\mathbb{K}$ can be defined as follows:

$$ICR_6(\mathbb{K}) = IC_6(R_{\mathbb{K}}) = -\log \frac{V(R_{\mathbb{K}})}{\int_{\mathbb{X}\times Y} \mathrm{d}\mathbf{x}\mathrm{d}y}, \tag{31}$$

$$ICR_7(\mathbb{K}) = IC_7(R_{\mathbb{K}}) = \left(\frac{1}{\ln 2}\right) \int_{\mathbb{X}\times Y} \frac{1 - \bar{R}_{\mathbb{K}}((\mathbf{x}, y))}{\mathbf{x} \times y} \mathrm{d}\mathbf{x}\mathrm{d}y. \tag{32}$$

Remark 4. $ICR_2(\mathbb{K})$, $ICR_4(\mathbb{K})$, $ICR_6(\mathbb{K})$ and $ICR_7(\mathbb{K})$ are obtained from $IC_2(\mathbb{K})$, $IC_4(\mathbb{K})$, $IC_6(\mathbb{K})$ and $IC_7(\mathbb{K})$ directly, while $ICR_1(\mathbb{K})$, $ICR_3(\mathbb{K})$ and $ICR_5(\mathbb{K})$ are different from $IC_1(R_{\mathbb{K}})$, $IC_3(R_{\mathbb{K}})$ and $IC_5(R_{\mathbb{K}})$. A fuzzy rule has an implication direction from antecedent to consequent, but we measure the information content of the fuzzy relation without direction in the definitions of $IC_1(R_{\mathbb{K}})$, $IC_3(R_{\mathbb{K}})$ and $IC_5(R_{\mathbb{K}})$. For $ICR_2(\mathbb{K})$, $ICR_4(\mathbb{K})$, $ICR_6(\mathbb{K})$ and $ICR_7(\mathbb{K})$, $R_{\mathbb{K}}$ is taken as a fuzzy set, and the direction has been concerned with the creation of the fuzzy relation.

Remark 5. When there is a single rule r in the rule base $\mathbb{K}$, Definitions 10 and 11 are obviously the definitions for the information content of the fuzzy rule r.

Definition 12. Let $\mathbb{K}$ be a fuzzy rule base, $\forall r \in \mathbb{K}$, the information importance of r in $\mathbb{K}$ is denoted as $IC\mathrm{Rim}_t^{\mathbb{K}}(r)$ and measured by

$$IC\mathrm{Rim}_t^{\mathbb{K}}(r) = ICR_t(\mathbb{K}) - ICR_t(\mathbb{K} \setminus \{r\}), \quad (t = 1, 2, \ldots, 8). \tag{33}$$

If $IC\mathrm{Rim}_t^{\mathbb{K}}(r) > 0$, rule r is defined as a effective rule in rule base $\mathbb{K}$;
If $IC\mathrm{Rim}_t^{\mathbb{K}}(r) = 0$, rule r is defined as a noneffective rule in rule base $\mathbb{K}$;
If $IC\mathrm{Rim}_t^{\mathbb{K}}(r) < 0$, rule r is defined as an inconsistent rule in rule base $\mathbb{K}$.

The information importance has been discussed in [20], which is a measure to determine the loss of information if rule r is omitted from the entire set of rule base $\mathbb{K}$.

In [19], the functions used to measure the information content are Gini-index

$$I_{\mathrm{Gini}}(\mathbb{K}) = 1 - \sum_{c=1}^{C} V_c(\mathbb{K})^2 \tag{34}$$

Table 2
The membership degree of $x \in X$ and $y \in Y$.

	−4	−3	−2	−1	0	1	2	3	4
PB	0	0	0	0	0	0.4	0.7	1	1
PS	0	0	0	0.4	0.7	1	0.7	0.4	0
ZO	0	0	0.4	0.7	1	0.7	0.4	0	0
NS	0	0.4	0.7	1	0.7	0.4	0	0	0
NB	1	1	0.7	0.4	0	0	0	0	0

and the fuzzy entropy function

$$I_{\text{Entropy}}(\mathbb{K}) = -\sum_{c=1}^{C} V_c(\mathbb{K}) \log V_c(\mathbb{K}), \tag{35}$$

where $V_c(\mathbb{K})$ indicates the volume of all rules in $\mathbb{K}$ which are assigned to class c.

Two of the most important properties of $I_{\text{Gini}}(\mathbb{K})$ and $I_{\text{Entropy}}(\mathbb{K})$ are

(1) $I_{\text{Gini}}(\mathbb{K}) = I_{\text{Entropy}}(\mathbb{K}) = 0$, when $V_c(\mathbb{K}) = 1$ for some c
(2) $I_{\text{Gini}}(\mathbb{K})$ and $I_{\text{Entropy}}(\mathbb{K})$ should be maximized when $V_c(\mathbb{K}) = 1/C$ for all c.

Thus, we think $I_{\text{Gini}}(\mathbb{K})$ and $I_{\text{Entropy}}(\mathbb{K})$ are more suitable for the measurement of the uncertainty or impurity than the information content of classification rule base $\mathbb{K}$.

The following two examples show how to use the measures of the information content of fuzzy rules and, by the way, how to use these measures to compare different implication operators.

Example 1. Let $X = \{-6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6\}$, $Y = \{-6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6\}$, *PB*, *PS*, *ZO*, *NS*, *NB* are fuzzy sets defined on X and Y. The membership function of them are shown in Table 2.

A rule r_1 is

r_1 : If Error is *NB*, Then Change is *PB*.

We choose O^5 as the implication operator, then r_1 induces a fuzzy relation R_{r_1},

$$R_{r_1} = \begin{pmatrix} 1 \\ 1 \\ 0.7 \\ 0.4 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0.4 & 0.7 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0.4 & 0.7 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0.4 & 0.7 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0.4 & 0.7 & 0.7 & 0.7 \\ 0 & 0 & 0 & 0 & 0 & 0.4 & 0.4 & 0.4 & 0.4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

All different values in $\{R_{r_1}(x, y) | x \in X, y \in Y\}$ can be designed to

$$\{1, 0.7, 0.4, 0\}$$

$$(R_{r_1})' = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0.4 & 0.7 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0.4 & 0.7 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0.4 & 0.7 & 0.7 & 0.7 \\ 0 & 0 & 0 & 0 & 0 & 0.4 & 0.4 & 0.4 & 0.4 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$$\begin{aligned} ICR_1(r_1) &= IC_1((R_{r_1})' \Downarrow X) = \sum_{i=1}^{9} \frac{|R_{r_1}(x_i)|}{|R'_{r_1}|} \log \frac{9}{|R_{r_1}(x_i)|} \\ &= \sum_{i=1}^{2} \frac{3.1}{55.3} \log \frac{9}{3.1} + \frac{2.5}{55.3} \log \frac{9}{2.5} + \frac{1.6}{55.3} \log \frac{9}{1.6} + \sum_{i=5}^{9} \frac{9}{55.3} \log \frac{9}{9} \\ &= 0.1724 + 0.0835 + 0.0721 = 0.3280 \end{aligned}$$

Table 3
The information content of $\mathbb{K}$ based on O^5.

	$ICR_1(\mathbb{K})$	$ICR_2(\mathbb{K})$	$ICR_2(\mathbb{K})$	$ICR_4(\mathbb{K})$	$ICR_5(\mathbb{K})$
O^5	1.1393	1.4301	0.8094	1.8717	0.9179

$$\begin{aligned}
ICR_2(r_1) &= IC_2(R_{r_1})\\
&= (1-0.7)\log\frac{81}{|(R_{r_1})_1|}+(0.7-0.4)\log\frac{81}{|(R_{r_1})_{0.7}|}+(0.4-0)\log\frac{81}{|(R_{r_1})_{0.4}|}+(0-0)\log\frac{81}{|(R_{r_1})_0|}\\
&= 0.3\log\frac{81}{4}+0.3\log\frac{81}{9}+0.4\log\frac{81}{16}+0\log\frac{81}{81}\\
&= 1.3020+0.9510+0.7020+0=2.955
\end{aligned}$$

$$H((R_{r_1})_1\Downarrow X)=H(((R_{r_1})_1)'\Downarrow X)=\frac{4}{67}\log\frac{9}{2}=0.1295,$$

$$H((R_{r_1})_{0.7}\Downarrow X)=H(((R_{r_1})_{0.7})'\Downarrow X)=\frac{9}{63}\log\frac{9}{3}=0.2264,$$

$$H((R_{r_1})_{0.4}\Downarrow X)=H(((R_{r_1})_{0.4})'\Downarrow X)=\frac{4}{67}\log\frac{9}{2}=0.3069,$$

$$H((R_{r_1})_0\Downarrow X)=H(((R_{r_1})_0)'\Downarrow X)=\frac{67}{67}\log\frac{9}{9}=0,$$

$$\begin{aligned}
ICR_3(r_1) &= 0.3\cdot H((R_{r_1})_1\Downarrow X)+0.3\cdot H((R_{r_1})_{0.7}\Downarrow X)+0.4\cdot H((R_{r_1})_{0.4}\Downarrow X)+0\cdot H((R_{r_1})_0\Downarrow X)\\
&= 0.3\times 0.1295+0.3\times 0.2264+0.4\times 0.3069+0\times 0=0.2295
\end{aligned}$$

$$\begin{aligned}
ICR_4(r_1) &= IC_4(R_{r_1})\\
&= \frac{1}{2.1}\log\frac{81}{|(R_{r_1})_1|}+\frac{0.7}{2.1}\log\frac{81}{|(R_{r_1})_{0.7}|}+\frac{0.4}{2.1}\log\frac{81}{|(R_{r_1})_{0.4}|}+\frac{0}{2.1}\log\frac{81}{|(R_{r_1})_0|}\\
&= \frac{1}{2.1}\log\frac{81}{4}+\frac{0.7}{2.1}\log\frac{81}{9}+\frac{0.4}{2.1}\log\frac{81}{16}+0\log\frac{81}{81}\\
&= 2.0666+1.0566+0.4457=3.5689
\end{aligned}$$

$$\begin{aligned}
IC_5(r_1) &= IC_5(R_{r_1})\\
&= \frac{1}{2.1}H((R_{r_1})_1\Downarrow X)+\frac{0.7}{2.1}H((R_{r_1})_{0.7}\Downarrow X)+\frac{0.4}{2.1}H((R_{r_1})_{0.4}\Downarrow X)+\frac{0}{2.1}H((R_{r_1})_0\Downarrow X)\\
&= 0.0617+0.0755+0.0585=0.1956.
\end{aligned}$$

If we have a rule base $\mathbb{K}$, $\mathbb{K}=\{r_1,r_2,r_3,r_4,r_5\}$,

r_2 : If Error is *NS*, Then Change is *PS*.
r_3 : If Error is *ZO*, Then Change is *ZO*.
r_4 : If Error is *PS*, Then Change is *NS*.
r_5 : If Error is *PB*, Then Change is *NB*.

The fuzzy relation induced by the rule base $\mathbb{K}$ is $R_{\mathbb{K}}=\bigcup_{i=1}^{5}R_{r_i}$, where R_{r_i} is induced by r_i.

$$R_{\mathbb{K}}=\bigcup_{i=1}^{5}R_{r_i}=\begin{array}{c}R\\-4\\-3\\-2\\-1\\0\\1\\2\\3\\4\end{array}\begin{array}{c}\begin{array}{ccccccccc}-4&-3&-2&-1&0&1&2&3&4\end{array}\\\left(\begin{array}{ccccccccc}
0&0&0&0&0&0.4&0.7&1&1\\
0&0&0&0.4&0.4&0.4&0.7&1&1\\
0&0&0.4&0.4&0.7&0.7&0.7&0.7&0.7\\
0&0.4&0.4&0.7&0.7&1&0.7&0.4&0.4\\
0&0.4&0.7&0.7&1&0.7&0.7&0.4&0\\
0.4&0.4&0.7&1&0.7&0.7&0.4&0.4&0\\
0.7&0.7&0.7&0.7&0.7&0.4&0.4&0&0\\
1&1&0.7&0.4&0.4&0.4&0&0&0\\
1&1&0.7&0.4&0&0&0&0&0
\end{array}\right)\end{array}.$$

Then, the information content of $\mathbb{K}$ can be evaluated by Definition 10, and the information importance of r_1 in $\mathbb{K}$ can be evaluated by Definition 12. The results are shown in Tables 3 and 4.

The results in Table 4 show us r_1 is a effective rule in $\mathbb{K}$ based on the measures $ICR_1(\star)$, $ICR_3(\star)$ and $ICR_5(\star)$, but a inconsistent rule in $\mathbb{K}$ by $ICR_2(\star)$ and $ICR_4(\star)$. This conflict is just the result of the different emphases of these measures, which have been discussed in Notice 3. In practice, we can choose a single measure to estimate the importance of rules, or aggregate the results with different weight. It must be noticed that $ICR_1(\star)$, $ICR_3(\star)$ and $ICR_5(\star)$ are more suitable for the estimation of the information importance of fuzzy rules in a rule base.

Table 4
$ICRim_t^{\mathbb{K}}(r_1)$ based on O^5.

	$ICRim_1^{\mathbb{K}}(r_1)$	$ICRim_2^{\mathbb{K}}(r_1)$	$ICRim_3^{\mathbb{K}}(r_1)$	$ICRim_4^{\mathbb{K}}(r_1)$	$ICRim_5^{\mathbb{K}}(r_1)$
O^5	0.1911	−0.3952	0.2885	−0.4769	0.3930

Table 5
The information content of r_1 based on different implication operators.

	$ICR_1(r_1)$	$ICR_2(r_1)$	$ICR_3(r_1)$	$ICR_4(r_1)$	$ICR_5(r_1)$
O^1	0.3122	0.4541	0.1850	0.5281	0.2003
O^2	0.0910	0.1154	0.0568	0.1674	0.0811
O^3	0.2280	0.3600	0.2342	0.4106	0.2683
O^4	0.2681	2.3600	0.2365	2.7229	0.2896
O^5	0.3280	2.955	0.2295	3.5689	0.1956
O^6	0.3109	3.3390	0.2326	3.6258	0.2020
O^7	0.3211	3.2717	0.2310	3.4133	0.2128

Based on different implication operators, we can obtain different fuzzy relations corresponding to r_1. For example, we can obtain $R_{r_1}^k$ by $O^k(k = 1, 2, 3, 4, 5, 6, 7)$ in Table 1. The information content of r_1 related to different $R_{r_1}^k$ are shown in Table 5.

Because $ICR_t(r_1),\ t = 1, \ldots, 5$, emphasize different sides of the information content of r_1, we can find that different $ICR_t(r_1)$ presents a different order of $O^i,\ i = 1, \ldots, 7$.

$$\begin{aligned}
&ICR_1(r_1) : O^5 > O^7 > O^1 > O^6 > O^4 > O^3 > O^2.\\
&ICR_2(r_1) : O^6 > O^7 > O^5 > O^4 > O^1 > O^3 > O^2.\\
&ICR_3(r_1) : O^6 > O^7 > O^5 > O^4 > O^1 > O^3 > O^2.\\
&ICR_4(r_1) : O^4 > O^3 > O^6 > O^7 > O^5 > O^1 > O^2.\\
&ICR_5(r_1) : O^4 > O^3 > O^7 > O^6 > O^1 > O^5 > O^2.
\end{aligned}$$

When we use the information content of r_1 as a criterion to choose a suitable implication operator, we can fix on a definition, $ICR_t(r_1)$, and use it to appraise the information power of the implication operators. For example, we choose the definition $ICR_1(r_1)$, then O^5 is the best implication operator which can make the rule r_1 represent the most information in classification and inference. When we make the choice, the fact must be noticed that $\{ICR_1(r_1), ICR_3(r_1), ICR_5(r_1)\}$ and $\{ICR_2(r_1), ICR_4(r_1)\}$ are two different classes of definition.

Sometimes, using some definition, $ICR_t(r_1)$, solely may lead to a unilateral result. We can integrate the function of the five definitions of the information content of r_1, and at this time, aggregation operator will be a good tool.

Example 2. Let $X = [-5, 5]$, $Y = [-1, 1]$, A, B are fuzzy sets defined on X and Y respectively.

$$A(x) = e^{-x^2}, \qquad B(y) = e^{-2(y-0.5)^2}.$$

For a rule "$r : A \to B$", the information content can be measured by Definition 10. The fuzzy relations induced by r with different implication operators in Table 1 are shown in Fig. 1, where "A and B" shows the fuzzy membership functions of A and B, "Ot" shows the fuzzy relation induced by r based on the implication operator O^t $(t = 1, \ldots, 7)$.

The information content of r based on O^t $(t = 1, 2, \ldots, 7)$ is shown in Table 6, and the information function of different implication operators can be ordered as follows:

$$O^6 > O^7 > O^5 > O^1 > O^2 > O^4 > O^3. \tag{36}$$

In fuzzy control, when the error is small, we should choose the fuzzy rule base to be more smooth and stable. But how to measure the degree of smoothness and stableness of fuzzy rule base is a difficult problem. Some discussions have been done in [23], which focuses on the choices of $\vee$ and $\wedge$ with minimal uncertainty. Compared with [23], the information content measures proposed by us can be used to evaluate the stability and sensitivity of fuzzy implication operators in fuzzy control. The implication operator which makes the given fuzzy rule more informative is more sensitive, otherwise, the implication operator is more stable. In Example 2, suppose $\{r\}$ is the rule base of a fuzzy controller, when the error of the system is smaller and smaller, we should choose the implication operators from left to right of Eq. (36). Otherwise, the implication operators should be chosen from right to left of Eq. (36), so that the controller can be changed to be more and more sensitive.

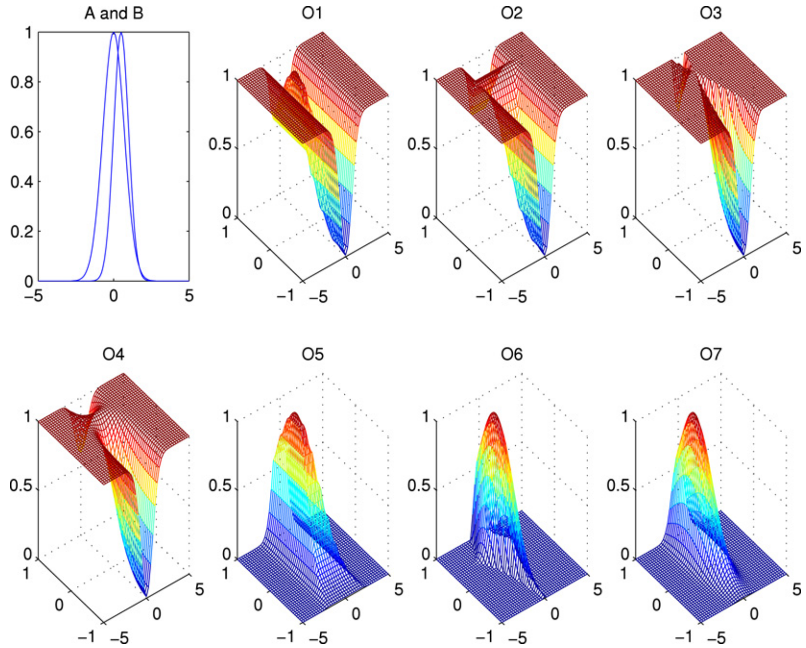

Fig. 1. The fuzzy relations induced by r based on O^t.

Table 6
The information content of r based on $O^t (t = 1, 2, \ldots, 7)$.

$ICR_7^{O^1}(r)$	$ICR_7^{O^2}(r)$	$ICR_7^{O^3}(r)$	$ICR_7^{O^4}(r)$	$ICR_7^{O^5}(r)$	$ICR_7^{O^6}(r)$	$ICR_7^{O^7}(r)$
0.1884	0.1568	0.0971	0.1265	3.1563	3.7518	3.422

4. Conclusion

Due to the importance of fuzzy rules in fuzzy modeling, fuzzy controllers and fuzzy expert systems, we try to propose some useful measures, which are different from confidence and coverage, for the estimation of fuzzy rules and rule bases. Since fuzzy rules can be fully captured by fuzzy relations, we discuss the information content of fuzzy relations at first. In this paper, five information content measures for fuzzy relations on discrete domains are proposed, and they can be partitioned to two groups with different emphases. One group emphasizes on the corresponding relations between domains (eg. X and Y), while the other emphasizes on the distinguishability of all elements in the Cartesian set of domains (eg. $X \times Y$). Then, measures about the information content measures of fuzzy relations on continuous domains and n-ary fuzzy relations are introduced, and these discussions make the work of this paper more integrated. Based on measures proposed for the information content of fuzzy relations, seven measures for the information content of fuzzy rule bases and rules are put forward. We can use these measures to choose rules with more information and do comparisons among different implication operators.

Fuzzy rule bases and fuzzy relations are widely used in different domains, and the estimation of them can lead to some useful algorithms. In the future, we will discuss the application of these measures in data mining (rule pruning), fuzzy decision tree learning (fuzzy relation estimation), and fuzzy control (variable implication operator in different processes of control).

References

[1] R.R. Yager, Entropy measures under similarity relations, International Journal of General Systems (1992) 20341–20358.
[2] E. Hernandez, J. Recasens, A reformulation of entropy in the presence of indistinguishability operators, Fuzzy Sets and Systems (2002) 128185–128196.
[3] D. Yu, et al., Uncertainty measures for fuzzy relations and their applications, Applied Soft Computing Journal (2006) doi:10.1016/j.asoc.2006.10.004.
[4] E.H. Ruspini, A new approach to clustering, Information and Control 15 (1969) 22–32.
[5] L.A. Zadeh, Probability measures of fuzzy events, Journal of Mathematical Analysis and Applications 23 (1968) 81–84.
[6] H. Tanaka, et al., A formulation of a fuzzy decision problem and its applications to an investment problem, Kybernetes 5 (1976) 25–30.
[7] D. Dumitrescu, Fuzzy measures and the entropy of fuzzy partitions, Journal of Mathematical Analysis and Applications 176 (1993) 359–373.
[8] R. Mesiar, J. Rybarik, Entropy of fuzzy partitions: A general model, Fuzzy Sets and Systems 99 (1998) 73–79.
[9] C. Bertoluzza, et al., Uncertainty measure on fuzzy partitions, Fuzzy Sets and Systems 142 (2004) 105–116.
[10] P. Matsakis, et al., Evaluation of fuzzy partitions, Remote Sensing of Environment 74 (2000) 516–533.
[11] G.M. Foody, Approaches for the production and evaluation of fuzzy land cover classifications from remotely sensed data, International Journal of Remote Sensing 17 (1996) 1317–1340.

[12] G.J. Klir, T.A. Folger, Fuzzy Sets, Uncertainty, and Information, Prentice Hall, USA, 1988.
[13] P. Ferreira, M. Reis, Fuzzy information on discrete and continuous domain: Approximation results, International Journal of General Systems 33 (5) (2004) 583–591.
[14] D. Hu, H.X. Li, The entropy of relations and a new approach for decision tree learning, Lecture Notes in Artifical Intelligence 3614 (2005) 378–388.
[15] T. Calvo, et al., A review of aggregation operators, Agop2001, Asturias, July 10–13, 2001.
[16] R.R. Yager, Measures of entropy and fuzziness related to aggregation operators, Information Sciences 82 (1995) 147–166.
[17] D. Dubois, et al., A note on quality measures for fuzzy association rules, in: Proceeding of the 10th International Fuzzy Systems Association World Congress on Fuzzy Sets and Systems, Istanbul, Turkey, in: LNAI, vol. 2715, 2003, pp. 346–353.
[18] Y.C. Hu, et al., Finding fuzzy classification rules using data mining techniques, Patter Recognition Letters 24 (2003) 509–519.
[19] M.R. Berthold, Mixed fuzzy rule formation, International Journal of Approximate Reasoning 32 (2003) 67–84.
[20] V. Cherkassky, F. Mulier, Learning from Data, John Wiley and Sons Inc., New York, 1998.
[21] D. Hand, J. Kok, M. Berthold (Eds.), Advances in Intelligent Data Analysis, in: Lecture Notes in Computer Science, vol. 1642, Springer, Berlin, 1999.
[22] R. Silipo, R.B. Michael, Input features impact on fuzzy decision processes, IEEE Transaction on Systems, Man, and Cybernetics, Part B: Cybernetics 30 (6) (2000) 821–834.
[23] A. Ramer, V. Kreinovich, Information complexity and fuzzy control. www.cs.utep.edu/vladik/1993/tr93-6.pdf.